Variable Step-Size Widely Linear Complex-Valued Affine Projection Algorithm and Performance Analysis

Long Shi, *Student Member, IEEE,* Haiquan Zhao, *Senior Member, IEEE,* Yuriy Zakharov, *Senior Member, IEEE* Badong Chen, *Senior Member, IEEE* and Yaoru Yang

***Abstract*—In this paper, a variable step-size widely linear complex-valued affine projection algorithm (VSS-WLCAPA) is proposed for processing noncircular signals. The variable step-size (VSS) is derived by minimizing the power of the augmented noise-free *a posteriori* error vector, which speeds up the convergence and reduces the steady-state misalignment. By exploiting the evolution of the covariance matrix of the weight error vector, we provide insight into the theoretical behavior of the VSS-WLCAPA algorithm. In the analysis, we take into account the dependency between the weight error vector and the noise vector, which is useful for accuracy of the theoretical prediction. To evaluate the mean step-size, the probability density function of the magnitude of the error is derived by employing polar coordinate transformation. Moreover, a special case when the projection order reduces to one is analysed in detail. The presented theoretical analysis is different from existing methodologies for analyzing affine projection algorithms due to the use of the Kronecker product. Simulation results for system identification scenarios demonstrate the merits of the proposed algorithm and verify the accuracy of the theoretical analysis. Wind prediction experiments support the superiority of the proposed VSS-WLCAPA as well.**

***Index Terms*—adaptive filter, *a posteriori* error, affine projection, covariance matrix, polar coordinate transformation, probability density function, variable step-size, widely linear**

I. Introduction

In the field of adaptive filtering, the least-mean-square (LMS) algorithm is widely considered as the fundamental algorithm owing to its simplicity [1], [2]. However, the conventional LMS algorithm experiences a gradient noise amplification problem. To overcome this shortcoming, the normalized LMS (NLMS) algorithm was introduced, which uses the squared Euclidean norm of input vector to normalize the step-size [3]. In real-world scenarios, e.g., the acoustic echo cancellation [4], [5], the input signals are highly correlated, which slows down the convergence of the LMS and NLMS algorithms [6].

To address the aforementioned problem, the affine projection algorithm (APA) has been proposed [7], [8]. Compared to the NLMS algorithm only relying on the current input vector, the APA updates the weight vector using current and past input vectors. To achieve fast convgence as well as low steady-state misalignment, the variable step-size (VSS) APA (VSS-APA) was proposed [9], where the VSS is derived by making the mean-square deviation (MSD) undergo the largest decrease from current iteration to next iteration. Thereafter, few more VSS strategies using distinct criterions were developed [10], [11]. Aimed at coping with other issues, several improved algorithms have been investigated [9], [12]–[16]. For example, the APA with evolving order (APA-EO) proposed in [16] is able to adaptively adjust the projection order to obtain better performance. Apart from the development of improved algorithms, the theoretical performance analysis of the APA family has also been given much attention [17]–[22]. In [17], the mean-square transient and steady-state performance of the APA was analyzed by using energy conservation arguments. For this analysis, the assumption of independence between the weight error vector and the noise vector was adopted to simplify the complicated derivation. It was reported in [19] that the weight error vector does depend on the noise vector. Taking into account this dependency, Kim *et al*. presented an improved mean-square error (MSE) analysis based on energy conservation arguments [20], which provides more accurate theoretical prediction of the algorithm performance. In [21], a novel analysis using the propagation model of the error covariance was put forward, which considers the cross-correlation between the weight error vector and the noise vector. Following the propagation model of the error covariance, the periodic APA (P-APA) [23] and the distributed filtered-x APA (DFxAPA) [24] were analyzed in detail.

In many practical applications, e.g., the frequency estimation in power systems [25], beamforming in communications [26], and wind prediction [27], the signals are complex-valued. Therefore, the research on the complex-valued adaptive filtering is of great interest. As an expansion of the LMS algorithm to the complex domain, the complex-valued LMS (CLMS) algorithm was proposed for circular signals whose second-order statistics are only described by the covariance matrix, while it is not applicable to noncircular signals which have non-zero pseudo-covariance matrice [28], [29]. To exploit the full information of noncircular signals, the widely linear CLMS (WL-CLMS) algorithm using the augmented complex-valued input vector has been proposed, achieving significant performance gain over the CLMS algorithm [30], [31]. To

L. Shi and H. Zhao are with the Key Laboratory of Magnetic Suspension Technology and Maglev Vehicle, Ministry of Education, and also with the School of Electrical Engineering, Southwest Jiaotong University, Chengdu 610031, People's Republic of China (e-mail: lshi@my.swjtu.edu.cn; hqzhao_swjtu@126.com), corresponding author: Haiquan Zhao.

Y. Zakharov and Yaoru Yang are with Department of Electronic Engineering, University of York, U.K. (e-mail: yury.zakharov@york.ac.uk, genuineyoung914@hotmail.com).

B. Chen is with the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an 710049, People's Republic of China (e-mail: chenbd@mail.xjtu.edu.cn).

circumvent the tradeoff caused by a fixed step-size in the WL-CLMS, the shrinkage WL-CLMS (SWL-CLMS) algorithm with VSS was proposed, which guarantees fast convergence and low steady-state misalignment [32]. With the widely linear model, the augmented complex-valued least-mean kurtosis (ACLMK) algorithm was designed to improve robustness against a wide range of noises [33]. In [34], Xia *et al*. proposed the augmented APA (AAPA), as well as presented the steady-state MSE analysis, where the dependency between the weight error vector and the noise vector was ignored for tractable analysis. However, this dependency can be important, as argued in [22]. The augmented complex-valued normalized subband adaptive filter (ACNSAF) algorithm was recently put forward in [35], which achieves performance improvement for colored input signals.

In this paper, we propose the VSS widely linear complex-valued APA (VSS-WLCAPA) algorithm to address the conflicting requirement between fast convergence and low steady-state misalignment. We investigate its performance via theoretical analysis, numerical simulations and wind prediction experiments. Owing to the use of the Kronecker product, the presented theoretical analysis is different from traditional methodologies used for analyzing the APA family. The main contributions of our work are summarized as follows:

1) The VSS is derived by minimizing the power of the augmented noise-free *a posteriori* error vector.
2) By considering the dependency between the weight error vector and the noise vector and applying some widely used assumptions, we arrive at the recursion of the covariance matrix of the weight error vector, which serves to perform the steady-state theoretical MSE prediction.
3) Since the variances of the real and imaginary parts of the error signal can be different, the magnitude of the error signal may not follow the Rayleigh distribution used in [36]. By employing polar coordinate transformation, we first derive the probability density function (PDF) of the magnitude of the error signal, which is then utilized to calculate the mean step-size.
4) The transient and steady-state performance for the VSS widely linear complex-valued NLMS (VSS-WLCNLMS) algorithm has been studied, which is a special case of the VSS-WLCAPA when the projection order reduces to one.

Simulations results for system identification scenarios show that the proposed VSS-WLCAPA outperforms the AAPA and ACNSAF for correlated input signals. The accuracy of the theoretical analysis is confirmed by simulation experiments implemented for white and correlated input signals. In addition, the wind prediction experiments implemented using the VSS-WLCAPA show more accurate prediction compared to using the AAPA.

The rest of this paper is organized as follows. In Section II, a review of the widely linear model and AAPA is presented, and then the derivation of the VSS-WLCAPA is given. In Section III, the theoretical behavior of the VSS-WLCAPA is analyzed in detail. In Section IV, simulations are carried out to test the proposed algorithm and verify the theoretical analysis. Finally, we draw conclusions in Section V.

Notation: Boldface letters denote vectors and matrices. Superscripts $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^{-1}$ are complex conjugate, transpose, Hermitian transpose, and matrix inverse, respectively. Symbols $\otimes$, $\max(\cdot)$, $|\cdot|$, and $\|\cdot\|^2$ stand for the Kronecker product, maximum, absolute operator for a scalar, and Euclidean norm of a vector, respectively. Operator $\mathrm{vec}(\cdot)$ stacks the matrix into a column. Symbols $E(\cdot)$, $\mathrm{Tr}(\cdot)$, $\det(\cdot)$, $\exp(\cdot)$ and $\mathrm{erf}(\cdot)$ denote the mathematical expectation, trace of a matrix, determinant of a matrix, exponential and error functions, respectively. Symbols $\mathbb{C}^{M\times 1}$ and $\mathbf{I_M}$ denote $M\times 1$ complex vector and identity matrix, respectively.

II. Widely Linear Model, AAPA and Proposed VSS-WLCAPA

A. Widely Linear Model

For a noncircular signal, its second-order statistics are described by the non-zero covariance matrix $\mathbf{R}$ and pseudo-covariance matrix $\mathbf{P}$ [29]. Considering a noncircular input vector $\mathbf{x}\in\mathbb{C}^{M\times 1}$, $\mathbf{R}$ and $\mathbf{P}$ are defined as [29]

$$\mathbf{R}=E\left[\mathbf{x}\mathbf{x}^H\right],\quad \mathbf{P}=E\left[\mathbf{x}\mathbf{x}^T\right]. \tag{1}$$

In order to exploit full second-order statistical information of $\mathbf{x}$, a widely linear model containing both $\mathbf{x}$ and $\mathbf{x}^*$ is given by [31]

$$y=\mathbf{h}^H\mathbf{x}+\mathbf{g}^H\mathbf{x}^*, \tag{2}$$

where y is the filter output, $\mathbf{h}$ and $\mathbf{g}$ are complex-valued coefficient vectors. Using the augmented complex-valued input vector $\mathbf{u}=[\mathbf{x}^T,\,\mathbf{x}^H]^T\in\mathbb{C}^{2M\times 1}$, the second-order statistics $\mathbf{R}$ and $\mathbf{P}$ are integrated into an augmented input covariance matrix

$$\mathbf{R}_u=E\left[\mathbf{u}\mathbf{u}^H\right]=\begin{bmatrix}\mathbf{R} & \mathbf{P}\\ \mathbf{P}^* & \mathbf{R}^*\end{bmatrix}. \tag{3}$$

In particular, the widely linear model reduces to the conventional linear model when the input signal is circular, i.e., $\mathbf{P}=\mathbf{0}$.

B. AAPA

Since the VSS-WLCAPA is developed based on the AAPA, we briefly introduce the AAPA. Consider the desired signal $d(k)$ originated from a widely linear model

$$d(k)=\mathbf{h}_o^H\mathbf{x}(k)+\mathbf{g}_o^H\mathbf{x}^*(k)+v(k), \tag{4}$$

where $\mathbf{h}_o\in\mathbb{C}^{M\times 1}$ and $\mathbf{g}_o\in\mathbb{C}^{M\times 1}$ are the unknown weight vectors, $\mathbf{x}(k)=[x(k),\cdots,x(k-M+1)]^T$ denotes the input vector, and $v(k)$ is the background noise with variance σ_v^2 at time instant k. The update equations of the AAPA are given by [34]

$$\begin{aligned}\mathbf{h}(k+1)=&\,\mathbf{h}(k)\\&+\mu\mathbf{X}^*(k)\left[\mathbf{X}^H(k)\mathbf{X}(k)+\mathbf{X}^T(k)\mathbf{X}^*(k)+\delta\mathbf{I}_P\right]^{-1}\mathbf{e}(k),\end{aligned} \tag{5}$$

$$\begin{aligned}\mathbf{g}(k+1)=&\,\mathbf{g}(k)\\&+\mu\mathbf{X}(k)\left[\mathbf{X}^H(k)\mathbf{X}(k)+\mathbf{X}^T(k)\mathbf{X}^*(k)+\delta\mathbf{I}_P\right]^{-1}\mathbf{e}(k),\end{aligned} \tag{6}$$

where $\mathbf{h}(k)$ and $\mathbf{g}(k)$ are weight vectors of the adaptive filter, μ is the step-size, $\mathbf{X}(k) = [\mathbf{x}(k), \cdots, \mathbf{x}(k-P+1)]$ denotes the $M \times P$ input matrix, P is the projection order, δ is a small positive constant to avoid numerical instability, $\mathbf{d}(k) = [d(k), \cdots, d(k-P+1)]^T$ is the desired vector, and $\mathbf{e}(k) = [e(k), \cdots, e(k-P+1)]^T$ represents the error vector calculated by

$$\mathbf{e}(k) = \mathbf{d}(k) - \mathbf{X}^T(k)\mathbf{h}(k) - \mathbf{X}^H(k)\mathbf{g}(k). \tag{7}$$

C. Proposed VSS-WLCAPA

Replacing μ in (5) and (6) with a time-varying step-size μ_k and denoting $\mathbf{S}(k) = \left[\mathbf{X}^H(k)\mathbf{X}(k) + \mathbf{X}^T(k)\mathbf{X}^*(k) + \delta\mathbf{I}_P\right]^{-1}$, we arrive at

$$\mathbf{h}(k+1) = \mathbf{h}(k) + \mu_k\mathbf{X}^*(k)\mathbf{S}(k)\mathbf{e}(k), \tag{8}$$

$$\mathbf{g}(k+1) = \mathbf{g}(k) + \mu_k\mathbf{X}(k)\mathbf{S}(k)\mathbf{e}(k). \tag{9}$$

By defining the weight error vectors $\tilde{\mathbf{h}}(k) = \mathbf{h}_o - \mathbf{h}(k)$ and $\tilde{\mathbf{g}}(k) = \mathbf{g}_o - \mathbf{g}(k)$, (8) and (9) are written as:

$$\begin{aligned}\tilde{\mathbf{h}}(k+1) = &\left\{\mathbf{I} - \mu_k\mathbf{X}^*(k)\mathbf{S}(k)\mathbf{X}^T(k)\right\}\tilde{\mathbf{h}}(k) \\ &- \mu_k\mathbf{X}^*(k)\mathbf{S}(k)\mathbf{X}^H(k)\tilde{\mathbf{g}}(k) - \mu_k\mathbf{X}^*(k)\mathbf{S}(k)\mathbf{v}(k),\end{aligned} \tag{10}$$

$$\begin{aligned}\tilde{\mathbf{g}}(k+1) = &\left\{\mathbf{I} - \mu_k\mathbf{X}(k)\mathbf{S}(k)\mathbf{X}^H(k)\right\}\tilde{\mathbf{g}}(k) \\ &- \mu_k\mathbf{X}(k)\mathbf{S}(k)\mathbf{X}^T(k)\tilde{\mathbf{h}}(k) - \mu_k\mathbf{X}(k)\mathbf{S}(k)\mathbf{v}(k).\end{aligned} \tag{11}$$

Combining (10) and (11) results in

$$\begin{aligned}\begin{bmatrix}\tilde{\mathbf{h}}(k+1)\\ \tilde{\mathbf{g}}(k+1)\end{bmatrix} = &\begin{bmatrix}\mathbf{I} - \mu_k\mathbf{X}^*(k)\mathbf{S}(k)\mathbf{X}^T(k) & \\ -\mu_k\mathbf{X}^*(k)\mathbf{S}(k)\mathbf{X}^H(k) & \end{bmatrix} \\ &\begin{matrix} & -\mu_k\mathbf{X}(k)\mathbf{S}(k)\mathbf{X}^T(k) \\ \mathbf{I} - \mu_k\mathbf{X}(k)\mathbf{S}(k)\mathbf{X}^H(k) & \end{matrix}\begin{bmatrix}\tilde{\mathbf{h}}(k)\\ \tilde{\mathbf{g}}(k)\end{bmatrix} \\ &- \mu_k\begin{bmatrix}\mathbf{X}^*(k)\mathbf{S}(k)\\ \mathbf{X}(k)\mathbf{S}(k)\end{bmatrix}\mathbf{v}(k),\end{aligned} \tag{12}$$

where $\mathbf{v}(k) = [v(k), \cdots, v(k-P+1)]$. Pre-multiplying (12) by the augmented input matrix $\mathbf{U}(k) = \left[\mathbf{X}^H(k), \mathbf{X}^T(k)\right]^T$ and ignoring the regularization term $\delta\mathbf{I}_P$ in $\mathbf{S}(k)$, we obtain

$$\mathbf{e}_p(k) = (1-\mu_k)\mathbf{e}_a(k) - \mu_k\mathbf{v}(k), \tag{13}$$

where $\mathbf{e}_p(k) = [e_p(k), \cdots, e_p(k-P+1)]^T$ is the noise-free *a posteriori* error vector given by

$$\mathbf{e}_p(k) = \mathbf{X}^T(k)\tilde{\mathbf{h}}(k+1) + \mathbf{X}^H(k)\tilde{\mathbf{g}}(k+1), \tag{14}$$

and $\mathbf{e}_a(k) = [e_a(k), \cdots, e_a(k-P+1)]^T$ is the noise-free *a priori* error vector given by

$$\mathbf{e}_a(k) = \mathbf{X}^T(k)\tilde{\mathbf{h}}(k) + \mathbf{X}^H(k)\tilde{\mathbf{g}}(k). \tag{15}$$

The Euclidean norm of (13) is then

$$\begin{aligned}\|\mathbf{e}_p(k)\|^2 = &(1-\mu_k)^2\|\mathbf{e}_a(k)\|^2 - \mu_k(1-\mu_k)\mathbf{e}_a^H(k)\mathbf{v}(k) \\ &- \mu_k(1-\mu_k)\mathbf{v}^H(k)\mathbf{e}_a(k) + \mu_k^2\|\mathbf{v}(k)\|^2.\end{aligned} \tag{16}$$

Invoking the widely used assumption of independence between $\mathbf{e}_a(k)$ and $\mathbf{v}(k)$ [17], it is easy to obtain $E\|\mathbf{e}(k)\|^2 = E\|\mathbf{e}_a(k)\|^2 + E\|\mathbf{v}(k)\|^2$. Then, taking the expectation in (16) yields

$$\begin{aligned}E\|\mathbf{e}_p(k)\|^2 &= (1-\mu_k)^2E\|\mathbf{e}_a(k)\|^2 + \mu_k^2E\|\mathbf{v}(k)\|^2 \\ &= (1-2\mu_k)E\|\mathbf{e}_a(k)\|^2 + \mu_k^2E\|\mathbf{e}(k)\|^2.\end{aligned} \tag{17}$$

The VSS is obtained by setting the derivative of (17) with respect to μ_k to 0. Thus, we arrive at

$$\mu_k = \frac{\sigma^2_{\mathbf{e}_a}(k)}{\sigma^2_{\mathbf{e}}(k)}, \tag{18}$$

where $\sigma^2_{\mathbf{e}}(k)$ and $\sigma^2_{\mathbf{e}_a}(k)$ are estimates of $E\|\mathbf{e}(k)\|^2$ and $E\|\mathbf{e}_a(k)\|^2$, respectively, which can be updated as:

$$\sigma^2_{\mathbf{e}}(k) = \alpha\sigma^2_{\mathbf{e}}(k-1) + (1-\alpha)\|\mathbf{e}(k)\|^2, \tag{19}$$

$$\sigma^2_{\mathbf{e}_a}(k) = \alpha\sigma^2_{\mathbf{e}_a}(k-1) + (1-\alpha)\|\hat{\mathbf{e}}_a(k)\|^2, \tag{20}$$

where

$$\hat{\mathbf{e}}_a(k) = \mathrm{sign}[\mathbf{e}(\mathrm{k})] \odot \hat{\mathbf{e}}(\mathrm{k}), \tag{21}$$

α is a forgetting factor ($0 < \alpha \leq 1$), $\mathrm{sign}[\mathbf{e}(\mathrm{k})]$ is a vector whose ith element is given by $\frac{e(i)}{|e(i)|}$, $\odot$ denotes the element-wise product, $\hat{\mathbf{e}}(k)$ is a vector whose ith element is calculated as $\max(|e(i)| - t, 0)$, and $t = \sqrt{\theta\sigma_v^2}$ is a threshold with θ being a constant, typically chosen in the interval $0 < \theta < 4$ [37]. We assume that the variance σ_v^2 of the background noise is known; if the noise variance is unknown, one can find its estimate using a recursive estimation method from [38], [39].

Practical consideration: In some practical applications, adaptive filtering algorithms adopt a small step-size to update their weight vectors [25], [27]. Since at the start of iterations, for our proposed algorithm, the power of noise-free a *priori* error $\sigma^2_{\mathbf{e}_a}(k)$ is close to the power of error $\sigma^2_{\mathbf{e}}(k)$, then from (18) it is found that the VSS starts from a large value (close to 1), which may lead to the algorithm divergence. Taking this into account, we assign a scale factor for the step-size to control its initial value, given by $\mu_k = \gamma_s\frac{\sigma^2_{\mathbf{e}_a}(k)}{\sigma^2_{\mathbf{e}}(k)}$, where γ_s is the scale factor chosen within $(0, 1]$.

III. Performance Analysis

In this section, we present the theoretical analysis of the VSS-WLCAPA performance. The scale factor γ_s for controlling the step-size is assumed to be 1. For compactness, the update equations (8) and (9) of the VSS-WLCAPA are written in the augmented form

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu_k\mathbf{U}(k)\mathbf{S}(k)\mathbf{e}(k), \tag{22}$$

where $\mathbf{w}(k) = \left[\mathbf{h}^T(k), \mathbf{g}^T(k)\right]^T$ is the augmented weight vector. In the subsequent analysis, the regularization term $\delta\mathbf{I}_P$ in $\mathbf{S}(k)$ is omitted.

To keep the analysis tractable, we make the following assumptions:

Assumption 1: The background noise $v(k)$ is a zero-mean noncircular white Gaussian sequence, which is independent of the input vector $\mathbf{x}(k)$ [1], [2].

Assumption 2: The step-size μ_k is independent of the input vector $\mathbf{x}(k)$ and the weight vector $\mathbf{w}(k)$ [3], [9], [40].

Assumption 3: The weight vector $\mathbf{w}(k)$ is independent of $\mathbf{U}(k)\mathbf{S}(k)\mathbf{U}^H(k)$ [8], [16].

Assumption 4: The noise-free *a priori* error $e_a(k)$ follows the zero-mean Gaussian distribution [37], [41]–[43].

Assumption 2 is commonly applied to analyze the VSS adaptive filtering algorithms by considering that the step-size varies slowly (α is close to 1). Assumption 3 is widely used in the APA family analysis [8]. Assumption 4 is approximately true when the filter length is large [37], [42].

Define the augmented weight error vector $\tilde{\mathbf{w}}(k) = \mathbf{w}_o - \mathbf{w}(k)$, where $\mathbf{w}_o = \left[\mathbf{h}_o^T, \mathbf{g}_o^T\right]^T$ is the augmented unknown system vector. From (22), we obtain

$$\begin{aligned}\tilde{\mathbf{w}}(k+1) &= \tilde{\mathbf{w}}(k) - \mu_k \mathbf{U}(k)\mathbf{S}(k)\mathbf{e}(k) \\ &= \{\mathbf{I} - \mu_k \mathbf{U}(k)\mathbf{S}(k)\mathbf{U}(k)\}\tilde{\mathbf{w}}(k) - \mu_k \mathbf{U}(k)\mathbf{S}(k)\mathbf{v}(k).\end{aligned} \tag{23}$$

Post-multiplying (23) by $\mathbf{w}^H(k+1)$, we arrive at

$$\begin{aligned}\mathbf{Q}(k+1) = \mathbf{Q}(k) &- \mu_k \mathbf{Q}(k)\mathbf{U}(k)\mathbf{S}(k)\mathbf{U}^H(k) \\ &- \mu_k \mathbf{U}(k)\mathbf{S}(k)\mathbf{U}^H(k)\mathbf{Q}(k) \\ &+ \mu_k^2 \mathbf{U}(k)\mathbf{S}(k)\mathbf{U}^H(k)\mathbf{Q}(k)\mathbf{U}(k)\mathbf{S}(k)\mathbf{U}^H(k) \\ &+ \mu_k^2 \mathbf{U}(k)\mathbf{S}(k)\mathbf{v}(k)\mathbf{v}^H(k)\mathbf{S}(k)\mathbf{U}^H(k) \\ &- \mu_k \tilde{\mathbf{w}}(k)\mathbf{v}^H(k)\mathbf{S}(k)\mathbf{U}^H(k) \\ &- \mu_k \mathbf{U}(k)\mathbf{S}(k)\mathbf{v}(k)\tilde{\mathbf{w}}^H(k) \\ &+ \mu_k^2 \mathbf{U}(k)\mathbf{S}(k)\mathbf{U}^H(k)\tilde{\mathbf{w}}(k)\mathbf{v}^H(k)\mathbf{S}(k)\mathbf{U}^H(k) \\ &+ \mu_k^2 \mathbf{U}(k)\mathbf{S}(k)\mathbf{v}(k)\tilde{\mathbf{w}}^H(k)\mathbf{U}(k)\mathbf{S}(k)\mathbf{U}^H(k),\end{aligned} \tag{24}$$

where $\mathbf{Q}(k) = \tilde{\mathbf{w}}(k)\tilde{\mathbf{w}}^H(k)$.

For arbitrary matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ of compatible dimensions, $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^{\mathrm{T}} \otimes \mathbf{A})\mathrm{vec}(\mathbf{B})$ [1], [44]. Applying this operation to (24) gives rise to

$$\begin{aligned}\mathrm{vec}(\mathbf{Q}(k+1)) = \mathrm{vec}(\mathbf{Q}(k)) &\\ - \mu_k[(\mathbf{U}(k)\mathbf{S}(k)\mathbf{U}^H(k))^T \otimes \mathbf{I}]\mathrm{vec}(\mathbf{Q}(k)) &\\ - \mu_k[\mathbf{I} \otimes (\mathbf{U}(k)\mathbf{S}(k)\mathbf{U}^H(k))]\mathrm{vec}(\mathbf{Q}(k)) &\\ + \mu_k^2[(\mathbf{U}(k)\mathbf{S}(k)\mathbf{U}^H(k))^T \otimes (\mathbf{U}(k)\mathbf{S}(k)\mathbf{U}^H(k))]\mathrm{vec}(\mathbf{Q}(k)) &\\ + \mu_k^2 \mathrm{vec}[\mathbf{U}(k)\mathbf{S}(k)\mathbf{v}(k)\mathbf{v}^H(k)\mathbf{S}(k)\mathbf{U}^H(k)] &\\ - \mu_k \mathrm{vec}[\tilde{\mathbf{w}}(k)\mathbf{v}^H(k)\mathbf{S}(k)\mathbf{U}^H(k)] &\\ - \mu_k \mathrm{vec}[\mathbf{U}(k)\mathbf{S}(k)\mathbf{v}(k)\tilde{\mathbf{w}}^H(k)] &\\ + \mu_k^2 \mathrm{vec}[\mathbf{U}(k)\mathbf{S}(k)\mathbf{U}^H(k)\tilde{\mathbf{w}}(k)\mathbf{v}^H(k)\mathbf{S}(k)\mathbf{U}^H(k)] &\\ + \mu_k^2 \mathrm{vec}[\mathbf{U}(k)\mathbf{S}(k)\mathbf{v}(k)\tilde{\mathbf{w}}^H(k)\mathbf{U}(k)\mathbf{S}(k)\mathbf{U}^H(k)]. &\end{aligned} \tag{25}$$

By taking the expectation of (25) and invoking Assumptions 1 and 3, we obtain

$$\begin{aligned}\mathrm{vec}[E(\mathbf{Q}(k+1))] &= \mathrm{vec}[E(\mathbf{Q}(k))] \\ &- E\{\mu_k[(\mathbf{U}(k)\mathbf{S}(k)\mathbf{U}^H(k))^T \otimes \mathbf{I}]\}\mathrm{vec}[E(\mathbf{Q}(k))] \\ &- E\{\mu_k[\mathbf{I} \otimes (\mathbf{U}(k)\mathbf{S}(k)\mathbf{U}^H(k))]\}\mathrm{vec}[E(\mathbf{Q}(k))] \\ &+ E\{\mu_k^2[(\mathbf{U}(k)\mathbf{S}(k)\mathbf{U}^H(k))^T \otimes (\mathbf{U}(k)\mathbf{S}(k)\mathbf{U}^H(k))]\} \\ &\times \mathrm{vec}[E(\mathbf{Q}(k))] \\ &+ \sigma_v^2 E\{\mu_k^2 \mathrm{vec}[\mathbf{U}(k)\mathbf{S}^2(k)\mathbf{U}^H(k)]\} \\ &- E\{\mu_k \mathrm{vec}[\tilde{\mathbf{w}}(k)\mathbf{v}^H(k)\mathbf{S}(k)\mathbf{U}^H(k)]\} \\ &- E\{\mu_k \mathrm{vec}[\mathbf{U}(k)\mathbf{S}(k)\mathbf{v}(k)\tilde{\mathbf{w}}^H(k)]\} \\ &+ E\{\mu_k^2 \mathrm{vec}[\mathbf{U}(k)\mathbf{S}(k)\mathbf{U}^H(k)\tilde{\mathbf{w}}(k)\mathbf{v}^H(k)\mathbf{S}(k)\mathbf{U}^H(k)]\} \\ &+ E\{\mu_k^2 \mathrm{vec}[\mathbf{U}(k)\mathbf{S}(k)\mathbf{v}(k)\tilde{\mathbf{w}}^H(k)\mathbf{U}(k)\mathbf{S}(k)\mathbf{U}^H(k)]\}.\end{aligned} \tag{26}$$

$E\{\mu_k[(\mathbf{U}(k)\mathbf{S}(k)\mathbf{U}^H(k))^T \otimes \mathbf{I}]\}$ in (26) can be written as:

$$\begin{aligned}&E\{\mu_k[(\mathbf{U}(k)\mathbf{S}(k)\mathbf{U}^H(k))^T \otimes \mathbf{I}]\} \\ &\quad = E(\mu_k)E[(\mathbf{U}(k)\mathbf{S}(k)\mathbf{U}^H(k))^T \otimes \mathbf{I}] \\ &\qquad + E\{[\mu_k - E(\mu_k)][(\mathbf{U}(k)\mathbf{S}(k)\mathbf{U}^H(k))^T \otimes \mathbf{I}]\}.\end{aligned} \tag{27}$$

With Assumption 2, the last term in (27) is zero, which means $\mu_k - E(\mu_k) \approx 0$. In other words, the step-size μ_k is very close to the mean step-size $E(\mu_k)$ [43], [45], [46], Furthermore, based on this fact, its variance can be considered small, thus

$$E(\mu_k^2) \approx [E(\mu_k)]^2. \tag{28}$$

Applying Assumption 2 and (28) to (26) results in

$$\begin{aligned}\mathrm{vec}[E(\mathbf{Q}(k+1))] =& \mathrm{vec}[E(\mathbf{Q}(k))] - E(\mu_k)\mathbf{\Phi}_1 \mathrm{vec}[E(\mathbf{Q}(k))] \\ &+ [E(\mu_k)]^2 \mathbf{\Phi}_2 \mathrm{vec}[E(\mathbf{Q}(k))] + [E(\mu_k)]^2 \sigma_v^2 \mathrm{vec}(\mathbf{\Phi}_3) \\ &- E(\mu_k)\mathrm{vec}(\mathbf{\Phi}_4) - E(\mu_k)\mathrm{vec}(\mathbf{\Phi}_4^H) \\ &+ [E(\mu_k)]^2 \mathrm{vec}(\mathbf{\Phi}_5) + [E(\mu_k)]^2 \mathrm{vec}(\mathbf{\Phi}_5^H),\end{aligned} \tag{29}$$

where

$$\mathbf{\Phi}_1 = E[\mathbf{U}(k)\mathbf{S}(k)\mathbf{U}^H(k)]^T \otimes \mathbf{I} + \mathbf{I} \otimes E[\mathbf{U}(k)\mathbf{S}(k)\mathbf{U}^H(k)], \tag{30}$$

$$\mathbf{\Phi}_2 = E\left\{[\mathbf{U}(k)\mathbf{S}(k)\mathbf{U}^H(k)]^T \otimes [\mathbf{U}(k)\mathbf{S}(k)\mathbf{U}^H(k)]\right\}, \tag{31}$$

$$\mathbf{\Phi}_3 = E[\mathbf{U}(k)\mathbf{S}^2(k)\mathbf{U}^H(k)], \tag{32}$$

$$\mathbf{\Phi}_4 = E[\tilde{\mathbf{w}}(k)\mathbf{v}^H(k)\mathbf{S}(k)\mathbf{U}^H(k)], \tag{33}$$

and

$$\mathbf{\Phi}_5 = E[\mathbf{U}(k)\mathbf{S}(k)\mathbf{U}^H(k)\tilde{\mathbf{w}}(k)\mathbf{v}^H(k)\mathbf{S}(k)\mathbf{U}^H(k)]. \tag{34}$$

In the traditional analysis of the APA family, the assumption of independence between the weight error vector $\tilde{\mathbf{w}}(k)$ and the noise vector $\mathbf{v}(k)$ is used to simplify the analysis [12], [17]. However, in reality, $\tilde{\mathbf{w}}(k)$ is dependent on the past noise vectors $\mathbf{v}(k-1), \mathbf{v}(k-2), \cdots, \mathbf{v}(k-P)$ [19], [21]. We have found that if this is not taken into account, the analysis may result in incorrect prediction. In this paper, we take into account their dependency so that the quantities $\mathbf{\Phi}_4$ and $\mathbf{\Phi}_5$ are

not neglected. From (23), the relationship between the weight error vector $\tilde{\mathbf{w}}(k)$ and the past noise vectors is given by

$$\begin{aligned}\tilde{\mathbf{w}}(k) = &\prod_{j=1}^{P}\left[\mathbf{I}-\mu_{k-j}\mathbf{U}(k-j)\mathbf{S}(k-j)\mathbf{U}^H(k-j)\right]\\ &\times\tilde{\mathbf{w}}(k-P)\\ -\sum_{j=1}^{P}&\Bigg(\prod_{p=1}^{j}\Big[\mathbf{I}-\mu_{k-p+1}\mathbf{U}(k-p+1)\mathbf{S}(k-p+1)\\ &\mathbf{U}^H(k-p+1)\Big]\Bigg)\mu_{k-j}\mathbf{U}(k-j)\mathbf{S}(k-j)\mathbf{v}(k-j).\end{aligned} \tag{35}$$

Using (35) in (33) and (34), $\mathbf{\Phi}_4$ and $\mathbf{\Phi}_5$ are calculated. The detailed computation is shown in Appendix A.

A. Steady-state Performance of the VSS-WLCAPA

We now consider the steady-state performance obtained from (29). As $k\to\infty$, $\mathrm{vec}[E(\mathbf{Q}(k+1))]=\mathrm{vec}[E(\mathbf{Q}(k))]$ holds. Thus, (29) is rearranged as

$$\begin{aligned}\mathrm{vec}[E(\mathbf{Q}(\infty))] = &E(\mu_\infty)\sigma_v^2[\mathbf{\Phi}_1-E(\mu_\infty)\mathbf{\Phi}_2]^{-1}\mathrm{vec}[\mathbf{\Phi}_3]\\ &-[\mathbf{\Phi}_1-E(\mu_\infty)\mathbf{\Phi}_2]^{-1}\mathrm{vec}[\mathbf{\Phi}_4+\mathbf{\Phi}_4^H]\\ &+E(\mu_\infty)[\mathbf{\Phi}_1-E(\mu_\infty)\mathbf{\Phi}_2]^{-1}\mathrm{vec}[\mathbf{\Phi}_5+\mathbf{\Phi}_5^H].\end{aligned} \tag{36}$$

To compute $\mathrm{vec}[E(\mathbf{Q}(\infty))]$, we need to find $E(\mu_\infty)$. For our further steps, the following approximations are required:

Approximation 1: In the steady-state, the excess mean-square error (EMSE) is much smaller than the noise variance σ_v^2, so that $E|e(k)|^2\approx\sigma_v^2$ [2], [42].

Approximation 2: For the VSS in the form $\mu_k=\frac{\delta_1}{\delta_2}$, the expectation of the ratio can be approximated by the ratio of their expectations, i.e., $E\left\{\frac{\delta_1}{\delta_2}\right\}\approx\frac{E(\delta_1)}{E(\delta_2)}$ [45], [47].

Approximation 3: In the steady-state, $\frac{E\|\hat{\mathbf{e}}_a(k)\|^2}{E\|\mathbf{e}(k)\|^2}\approx\frac{E|\hat{e}_a(k)|^2}{E|e(k)|^2}$ holds.

Approximation 1 is widely used for analyzing the steady-state behavior of adaptive filtering algorithms. Approximation 2 has been successfully employed to analyze the VSS algorithms. We also illustrate the validity of this approximation in Appendix B. Approximation 3 is valid because in the steady-state $E|\hat{e}_a(k)|^2=E|\hat{e}_a(k-1)|^2=\cdots=E|\hat{e}_a(k-P+1)|^2$ and $E|e(k)|^2=E|e(k-1)|^2=\cdots=E|e(k-P+1)|^2$.

Recalling (18)-(20) and using Approximation 2, the steady-state mean step-size $E(\mu_\infty)$ is given by

$$E(\mu_\infty)=\frac{E\|\hat{\mathbf{e}}_a(\infty)\|^2}{E\|\mathbf{e}(\infty)\|^2}. \tag{37}$$

Furthermore, applying Approximation 3 to (37), we arrive at

$$E(\mu_\infty)=\frac{E|\hat{e}_a(\infty)|^2}{E|e(\infty)|^2}. \tag{38}$$

In (38), $E|e(\infty)|^2\approx\sigma_v^2$ according to Approximation 1, whereas the difficulty is in the calculation of $E|\hat{e}_a(\infty)|^2$. From (21), $\hat{e}_a(k)$ is expressed as

$$\hat{e}_a(k)=\mathrm{sign}[e(k)]\max(|e(k)|-t,0). \tag{39}$$

From (39), we obtain

$$E|\hat{e}_a(k)|^2=E\{[\max(|e(k)|-t)]^2\}. \tag{40}$$

To evaluate the expectation in (40), we need to know the PDF of $|e(k)|$. With Assumption 1 and Assumption 4, considering $e(k)=e_a(k)+v(k)$, it is apparent that $e(k)$ obeys the zero-mean Gaussian distribution. However, the variances of the real and imaginary parts can be different. Therefore, we cannot use results from [36], and we need to find the PDF of $|e(k)|$ in the following. Using polar coordinate transformation, the distribution of magnitude $|e(k)|$ is derived in Appendix C.

Let $r=|e(k)|$, the quantity $E|\hat{e}_a(k)|^2$ is then calculated as:

$$E|\hat{e}_a(k)|^2=\int_t^\infty(r-t)^2\frac{2r}{\sigma^2}\exp\left(-\frac{r^2}{\sigma^2}\right)\mathrm{d}r. \tag{41}$$

By taking the integral in (41), we obtain

$$E|\hat{e}_a(k)|^2=\frac{2}{\sigma^2}[\Omega_1-\Omega_2+\Omega_3], \tag{42}$$

where

$$\begin{aligned}\Omega_1&=\int_t^\infty r^3\exp\left(-\frac{r^2}{\sigma^2}\right)\mathrm{d}r\\ &=\frac{\sigma^2}{2}\left[t^2\exp\left(-\frac{t^2}{\sigma^2}\right)+\sigma^2\exp\left(-\frac{t^2}{\sigma^2}\right)\right],\end{aligned} \tag{43}$$

$$\begin{aligned}\Omega_2&=2t\int_t^\infty r^2\exp\left(-\frac{r^2}{\sigma^2}\right)\mathrm{d}r\\ &=t\sigma^2\left[t\exp\left(-\frac{t^2}{\sigma^2}\right)-\frac{\sqrt{\pi}\sigma\left(\mathrm{erf}\left(\frac{t}{\sigma}\right)-1\right)}{2}\right],\end{aligned} \tag{44}$$

and

$$\begin{aligned}\Omega_3&=t^2\int_t^\infty r\exp\left(-\frac{r^2}{\sigma^2}\right)\mathrm{d}r\\ &=\frac{t^2\sigma^2}{2}\exp\left(-\frac{t^2}{\sigma^2}\right).\end{aligned} \tag{45}$$

In the steady-state, using Approximation 1, we have $\sigma^2=\sigma_v^2$. From (38)-(45), we can now obtain $E(\mu_\infty)$. Eventually, we calculate the steady-state MSE as:

$$\mathrm{MSE}(\infty)=\sigma_v^2+\mathrm{Tr}\left(\mathbf{R}_u^{2M}E[\mathbf{Q}(\infty)]\right), \tag{46}$$

where $\mathbf{R}_u^{2M}=E[\mathbf{u}(k)\mathbf{u}^H(k)]$ denotes the $2M$th-order augmented input covariance matrix with $\mathbf{u}(k)$ being the first column vector of the augmented input matrix $\mathbf{U}(k)$.

Furthermore, applying the property $\mathrm{Tr}(\mathbf{XY})=\mathrm{vec}(\mathbf{X}^T)^T\mathrm{vec}(\mathbf{Y})$ [44] to (46) leads to

$$\mathrm{MSE}(\infty)=\sigma_v^2+\mathrm{vec}\left[(\mathbf{R}_u^{2M})^T\right]^T\mathrm{vec}\left(E[\mathbf{Q}(\infty)]\right). \tag{47}$$

Remark 1: If the real and imaginary parts of the error have the same variances, the PDF of the magnitude $|e(k)|$ in (C.8) turns into the form given in our previous work [36]. Therefore, the PDF of $|e(k)|$ derived in this paper can be regarded as a generalized expression for a complex-valued error signal.

Remark 2: In the steady-state, replacing σ^2 in (42)-(45) with σ_v^2, after some algebra, we have

$$E|\hat{e}_a(\infty)|^2=\sigma_v^2\left(\exp(-\theta)+\sqrt{\pi\theta}\left[\mathrm{erf}(\sqrt{\theta})-1\right]\right). \tag{48}$$

Then, $E(\mu_\infty)$ in (38) becomes

$$E(\mu_\infty) = \exp(-\theta) + \sqrt{\pi\theta}\left[\mathrm{erf}(\sqrt{\theta}) - 1\right]. \quad (49)$$

It is seen from (49) that the expectation of the steady-state step-size $E(\mu_\infty)$ only depends on the parameter θ; it does not depend on the input signal and noise variances.

Remark 3: With (49), we can obtain the steady-state step-size $E(\mu_\infty)$, which is then used to (36) to obtain $\mathrm{vec}[E(\mathbf{Q}(\infty))]$. The steady-state MSE is finally calculated using (47).

B. Transient and Steady-state Performance of the VSS-WLCNLMS

We next present a special case for the analysis: $P = 1$. Now, the proposed algorithm transforms into the VSS-WLCNLMS algorithm. In this case, $\mathbf{\Phi}_1$, $\mathbf{\Phi}_2$ and $\mathbf{\Phi}_3$ reduce to

$$\mathbf{\Phi}_1 = E[\mathbf{u}(k)s(k)\mathbf{u}^H(k)]^T \otimes \mathbf{I} + \mathbf{I} \otimes E[\mathbf{u}(k)s(k)\mathbf{u}^H(k)], \quad (50)$$

$$\mathbf{\Phi}_2 = E\left\{[\mathbf{u}(k)s(k)\mathbf{u}^H(k)]^T \otimes [\mathbf{u}(k)s(k)\mathbf{u}^H(k)]\right\}, \quad (51)$$

$$\mathbf{\Phi}_3 = E[\mathbf{u}(k)s^2(k)\mathbf{u}^H(k)], \quad (52)$$

where

$$s(k) = [\mathbf{x}^H(k)\mathbf{x}(k) + \mathbf{x}^T(k)\mathbf{x}^*(k)]^{-1}. \quad (53)$$

The quantities $\mathbf{\Phi}_4$ and $\mathbf{\Phi}_5$ are zeros due to the independency between the weight error vector $\tilde{\mathbf{w}}(k)$ and the noise $v(k)$ [43], [45]. Therefore, (29) takes the form

$$\begin{aligned}\mathrm{vec}[E(\mathbf{Q}(k+1))] =& \mathrm{vec}[E(\mathbf{Q}(k))] - E(\mu_k)\mathbf{\Phi}_1\mathrm{vec}[E(\mathbf{Q}(k))] \\ &+[E(\mu_k)]^2\mathbf{\Phi}_2\mathrm{vec}[E(\mathbf{Q}(k))] + [E(\mu_k)]^2\sigma_v^2\mathrm{vec}(\mathbf{\Phi}_3).\end{aligned} \quad (54)$$

When $P = 1$, the VSS in (18) becomes

$$\mu_k = \frac{\sigma_{e_a}^2(k)}{\sigma_e^2(k)}, \quad (55)$$

where

$$\sigma_e^2(k) = \alpha\sigma_e^2(k-1) + (1-\alpha)|e(k)|^2, \quad (56)$$

$$\sigma_{e_a}^2(k) = \alpha\sigma_{e_a}^2(k-1) + (1-\alpha)|\hat{e}_a(k)|^2, \quad (57)$$

and $\hat{e}_a(k)$ is given in (39).

Following the procedures in (41)-(45), we can calculate the mean step-size $E(\mu_k)$ for (55), which is then used to predict the transient behavior of the algorithm. When calculating the quantity $E|\hat{e}_a(k)|^2$ in (41), $\sigma^2 = \sigma_v^2 + \mathrm{Tr}\left(\mathbf{R}_u^{2M}E[\mathbf{Q}(k)]\right)$ is used. From (54), we obtain the steady-state formula given by

$$\mathrm{vec}[E(\mathbf{Q}(\infty))] = E(\mu_\infty)\sigma_v^2[\mathbf{\Phi}_1 - E(\mu_\infty)\mathbf{\Phi}_2]^{-1}\mathrm{vec}[\mathbf{\Phi}_3]. \quad (58)$$

We will also be using the following relationship:

$$\begin{aligned}E[\mathbf{u}^H(k)\mathbf{u}(k)] &= E[\mathbf{x}^H(k)\mathbf{x}(k) + \mathbf{x}^T(k)\mathbf{x}^*(k)] \\ &= \mathrm{Tr}(\mathbf{R}_u^{2M}).\end{aligned} \quad (59)$$

If the length M is large enough, it can be shown that [19]

$$\mathbf{x}^H(k)\mathbf{x}(k) + \mathbf{x}^T(k)\mathbf{x}^*(k) \approx \mathrm{Tr}(\mathbf{R}_u^{2M}). \quad (60)$$

Applying (59) in (53), we obtain

$$s(k) = [\mathrm{Tr}(\mathbf{R}_u^{2M})]^{-1}. \quad (61)$$

Particularly, when the white input signal is considered, the quantities $\mathbf{\Phi}_1$, $\mathbf{\Phi}_2$ and $\mathbf{\Phi}_3$ in (50)-(52) can be further simplified. Since

$$\mathbf{R}_u^{2M} = \begin{bmatrix}(\mathbf{R}_x^M)^* & (\mathbf{P}_x^M)^* \\ \mathbf{P}_x^M & \mathbf{R}_x^M\end{bmatrix} = \begin{bmatrix}\sigma_x^2\mathbf{I}_M & \tilde{\sigma}_x^2\mathbf{I}_M \\ \tilde{\sigma}_x^2\mathbf{I}_M & \sigma_x^2\mathbf{I}_M\end{bmatrix}, \quad (62)$$

it is straightforward to see that $s(k) = \frac{1}{2M\sigma_x^2}$ and $\mathbf{\Phi}_1$ can be written as

$$\mathbf{\Phi}_1 = \frac{1}{2M\sigma_x^2}\left\{\begin{bmatrix}\sigma_x^2\mathbf{I}_M & \tilde{\sigma}_x^2\mathbf{I}_M \\ \tilde{\sigma}_x^2\mathbf{I}_M & \sigma_x^2\mathbf{I}_M\end{bmatrix}^T \otimes \mathbf{I} + \mathbf{I} \otimes \begin{bmatrix}\sigma_x^2\mathbf{I}_M & \tilde{\sigma}_x^2\mathbf{I}_M \\ \tilde{\sigma}_x^2\mathbf{I}_M & \sigma_x^2\mathbf{I}_M\end{bmatrix}\right\}, \quad (63)$$

where $\mathbf{R}_x^M = \sigma_x^2\mathbf{I}_M$ denotes the Mth-order input covariance matrix with variance $\sigma_x^2 = E[x(k)x^*(k)]$, and $\mathbf{P}_x^M = \tilde{\sigma}_x^2\mathbf{I}_M$ is the Mth-order pseudo-covariance matrix with complementary variance $\tilde{\sigma}_x^2 = E[x^2(k)]$. Again, substituting $s(k)$ to (51) results in

$$\mathbf{\Phi}_2 = \frac{1}{[2M\sigma_x^2]^2}E\left\{[\mathbf{u}(k)\mathbf{u}^H(k)]^T \otimes [\mathbf{u}(k)\mathbf{u}^H(k)]\right\}. \quad (64)$$

By invoking the Gaussian moment factorizing theorem [43], [46] for the fourth order moment in (64), we obtain

$$\begin{aligned}\mathbf{\Phi}_2 = \frac{1}{[2M\sigma_x^2]^2}\big\{&(\mathbf{R}_u^{2M})^T \otimes \mathbf{R}_u^{2M} + (\mathbf{P}_u^{2M})^H \otimes \mathbf{P}_u^{2M} \\ &+\mathrm{vec}(\mathbf{R}_u^{2M})(\mathrm{vec}[(\mathbf{R}_u^{2M})^T])^T\big\},\end{aligned} \quad (65)$$

where

$$\begin{aligned}\mathbf{P}_u^{2M} = E[\mathbf{u}(k)\mathbf{u}^T(k)] &= \begin{bmatrix}(\mathbf{P}_x^M)^* & (\mathbf{R}_x^M)^* \\ \mathbf{R}_x^M & \mathbf{P}_x^M\end{bmatrix} \\ &= \begin{bmatrix}\tilde{\sigma}_x^2\mathbf{I}_M & \sigma_x^2\mathbf{I}_M \\ \sigma_x^2\mathbf{I}_M & \tilde{\sigma}_x^2\mathbf{I}_M\end{bmatrix}\end{aligned} \quad (66)$$

is the $2M$th-order augmented pseudo-covariance matrix. Similarly, $\mathbf{\Phi}_3$ is reformulated as

$$\mathbf{\Phi}_3 = \frac{1}{[2M\sigma_x^2]^2}\begin{bmatrix}\sigma_x^2\mathbf{I}_M & \tilde{\sigma}_x^2\mathbf{I}_M \\ \tilde{\sigma}_x^2\mathbf{I}_M & \sigma_x^2\mathbf{I}_M\end{bmatrix}. \quad (67)$$

Remark 4: Based on (50)-(52) and the mean step-size $E(\mu_k)$, $\mathrm{vec}\left(E[\mathbf{Q}(k)]\right)$ in (54) can be recursively calculated. Then, the transient behavior of the VSS-WLCNLMS algorithm is predicted by using $\mathrm{MSE}(k) = \sigma_v^2 + \mathrm{vec}\left[(\mathbf{R}_u^{2M})^T\right]^T\mathrm{vec}\left(E[\mathbf{Q}(k)]\right)$. Combining (58) and (47), the steady-state MSE is obtained. When considering the case of white input signal, in (58) we use (63), (65) and (67) to calculate $\mathbf{\Phi}_1$, $\mathbf{\Phi}_2$ and $\mathbf{\Phi}_3$, respectively.

IV. Simulation

In this section, we will show the performance results of the proposed algorithm, and verify the accuracy of the theoretical prediction by numerical simulation. The MSD curves, i.e., $10\log10\|\mathbf{w}_o - \mathbf{w}(k)\|^2$, and the MSE curves, i.e., $10\log10|e(k)|^2$, are plotted to measure the algorithm performance and validate the theoretical analysis, respectively.

A. System Identification Scenarios

The unknown system vectors $\mathbf{h}_o$ and $\mathbf{g}_o$ are randomly generated from a zero-mean Gaussian distribution with unit norm, respectively. When evaluating the algorithm tracking capability, the unknown system vectors $\mathbf{h}_o$ and $\mathbf{g}_o$ change to $-\mathbf{h}_o$ and $-\mathbf{g}_o$ in the middle of iterations. The initial weights of the adaptive filter are set to zeros. The correlated input signal is an autoregressive first order (AR1) Gaussian process. The white Gaussian input is zero-mean non-circular with variance $\sigma_x^2 = 1$ and complementary variance $\tilde{\sigma}_x^2 = 0.5$. When investigating the proposed algorithm, the highly correlated input is considered for the VSS-WLCAPA with projection order higher than one, while the weakly correlated or white Gaussian input is considered for the VSS-WLCNLMS. The background noise is zero-mean white Gaussian. The reguarization δ is set to 10^{-6}, and the scale factor γ_s is set to 1. All results are obtained by averaging over 100 independent trials.

We first plot the MSD curves of the proposed VSS-WLCAPA with different θ and α, as shown in Figs. 1 and 2. As can be seen from Fig. 1, with the increase of θ, the steady-state misalignment of the VSS-WLCAPA is reduced until $\theta = 3$, and then increases. In Fig. 2, as α increases, the convergence of the VSS-WLCAPA slows down. After sudden change of the system vectors, the proposed algorithm maintains similar performance for the same θ and α. In the following experiments, we select $\alpha = 0.95$ and $\theta = 3$ for the proposed algorithm.

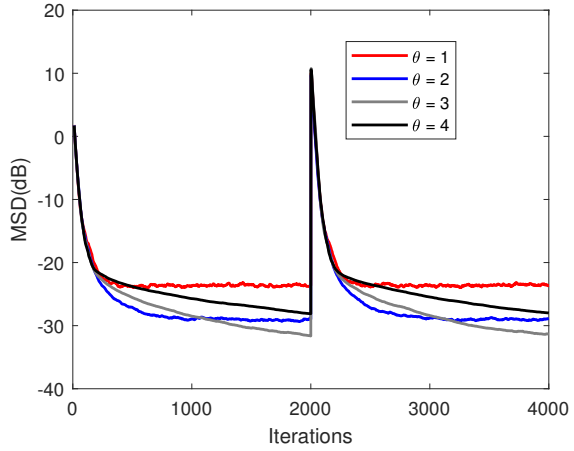


Fig. 1. MSD curves of the proposed algorithm against θ for the AR1(0.95) input signal. $P = 4, M = 8, \alpha = 0.95$ and $\sigma_v^2 = 0.01$.

Fig. 3 compares the proposed algorithm with the ACNSAF and AAPA for different P and M. The proposed VSS-WLCAPA outperforms the ACNSAF and AAPA; it provides faster convergence and lower steady-state misalignment. In addition, the proposed algorithm has better tracking capability after the change of the unknown system than the other algorithms. Fig. 4 shows the evolutions of the step-size of the proposed VSS-WLCAPA. It is seen that the VSS varies from a large value, then rapidly reduces to a small value. This evolution accounts for fast convergence at the start of

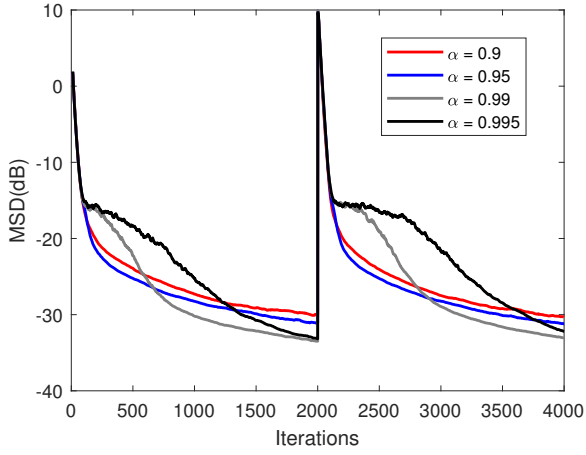


Fig. 2. MSD curves of the proposed algorithm against α for the AR1(0.95) input signal. $P = 4, M = 8, \theta = 3$ and $\sigma_v^2 = 0.01$.

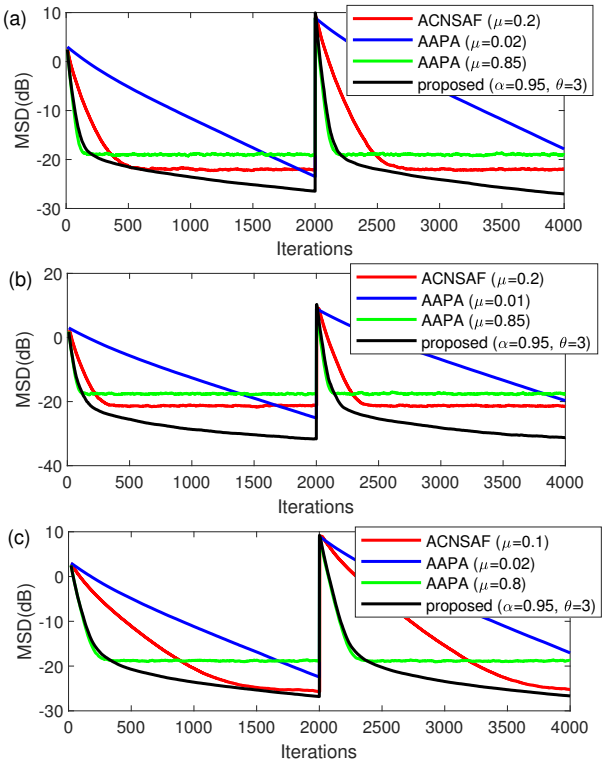


Fig. 3. Comparison of the ACNSAF, AAPA and proposed VSS-WLCAPA for the AR1(0.95) input signal, $\sigma_v^2 = 0.01$. (a) $N_{sub} = 2, P = 2, M = 8$ (b) $N_{sub} = 4, P = 4, M = 8$ (c) $N_{sub} = 4, P = 4, M = 16$, where N_{sub} denotes the number of subbands of the ACNSAF.

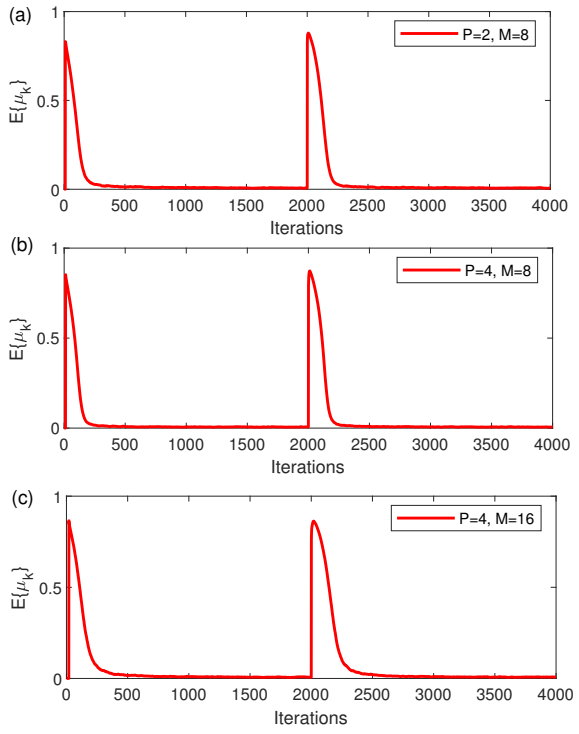


Fig. 4. Evolutions of the VSS of the proposed VSS-WLCAPA for the AR1(0.95) input signal, $\alpha = 0.95$, $\theta = 3$ and $\sigma_v^2 = 0.01$.

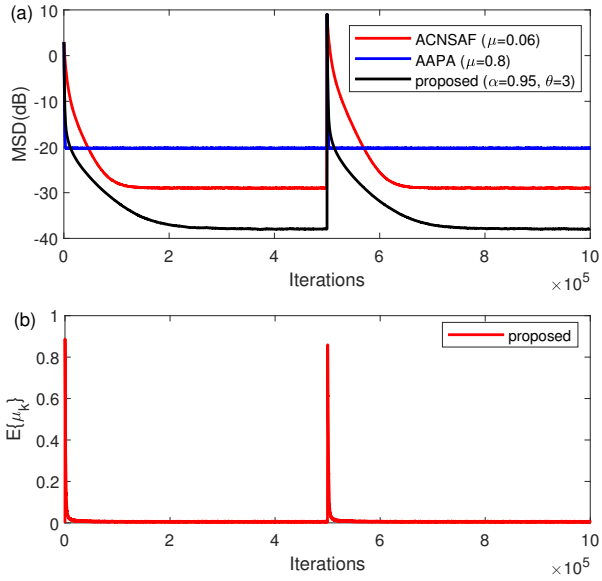


Fig. 5. Comparison of the ACNSAF, AAPA and proposed VSS-WLCAPA for the AR1(0.95) input signal, $\sigma_v^2 = 0.01$, $N_{sub} = 2, P = 2, M = 128$ (a) MSD curves of various algorithms (b) evolutions of the VSS of the proposed VSS-WLCAPA.

iterations and low steady-state misalignment in the steady-state. Moreover, the VSS shows much the same evolutionary trend after the change of the unknown system.

In Fig. 5, the filter length is increased to $M = 128$. It is evident from Fig. 5(a) that the proposed VSS-WLCAPA has better performance than the ACNSAF and AAPA. The VSS in Fig. 5(b) follows the trend of evolving from a large value to a small value.

We next validate the theoretical analysis. As discussed in Remark 2, the theoretical steady-state step-size $E(\mu_\infty)$ depends on θ. In Fig.6, we show the accuracy of the theoretical $E(\mu_\infty)$ calculated using (49) for different θ. It is seen that there is some deviation between the simulated and theoretical values when $\theta \leq 1$, while the prediction is very accurate for $\theta > 1$.

Fig. 7 shows the steady-state MSE of the proposed VSS-WLCAPA against θ. The theoretical MSE is calculated according to Remark 3. As can be seen in Fig. 7, the theoretical result has some deviation from the simulated result, and with the increase of θ, the deviation is reduced. This is because at higher θ, the theoretical step-size predicts the simulated step-size more accurately. It is also observed that there exists more significant deviation between the theoretical and simulated results for higher P ($P = 4, M = 8$). This is because of the limits of the approximation in (28) when the filter length is not far larger than the projection order.

We now validate the theoretical prediction of the MSE performance for $P = 1$. In Fig. 8, we plot the evolutions of the VSS for different θ and α. As can be seen, the theoretical step-sizes are in a good agreement with the simulation results.

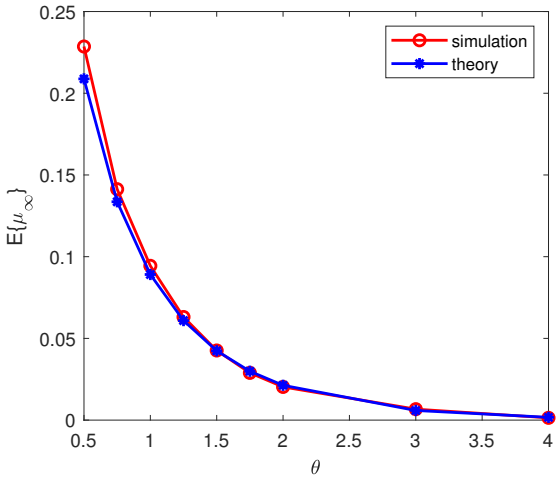


Fig. 6. Steady-state step-size $E\{\mu_\infty\}$ against θ in the proposed algorithm for the AR1(0.95) input signal. $P = 2, M = 8, \alpha = 0.95$ and $\sigma_v^2 = 0.01$.

Fig. 9 shows the transient MSE performance of the proposed algorithm for different σ_v^2. The theoretical result is calculated according to Remark 4. When $\sigma_v^2 = 0.1$ and $\sigma_v^2 = 0.01$, the theoretical results match well with the simulation results. Only for a low noise variance ($\sigma_v^2 = 0.001$), the theoretical prediction deviates from the simulation results.

We next verify the theoretical analysis for white input signal. As can be seen from Fig. 10, when $\alpha = 0.99, \theta = 3$, the simulated VSS is accurately approximated by the theoretical VSS, while when $\alpha = 0.995, \theta = 4$, there exists some deviation between them. This is due to the limited accuracy of the approximation in (38) and (60). Fig. 11 shows the

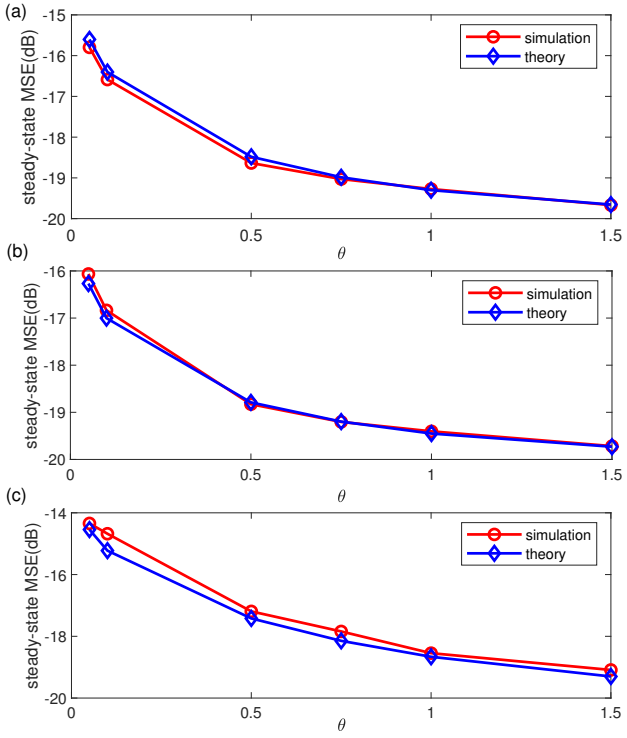


Fig. 7. Steady-state MSE of the proposed algorithm against θ for the AR1(0.95) input signal, $\alpha = 0.95$, and $\sigma_v^2 = 0.01$. (a) $P = 2, M = 8$ (b) $P = 2, M = 16$ (c) $P = 4, M = 8$.

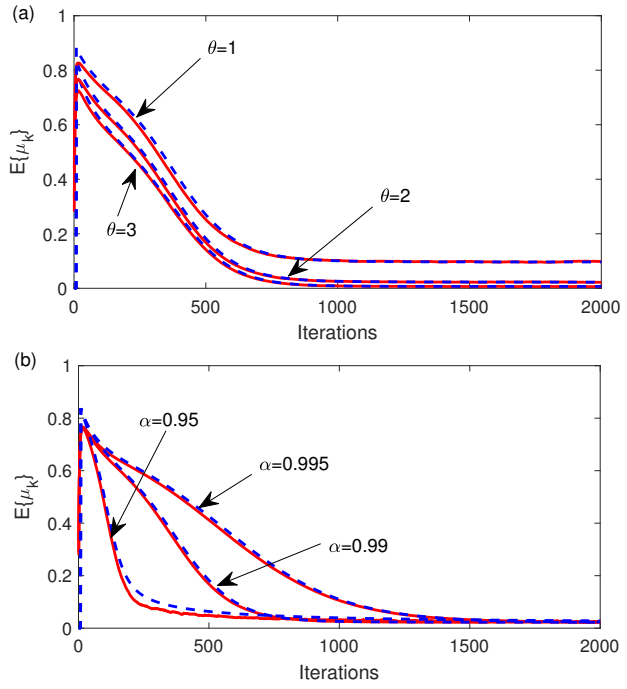


Fig. 8. Evolutions of the proposed VSS for the AR1(0.5) input signal, $P = 1, M = 8$ and $\sigma_v^2 = 0.01$. (a) $\alpha = 0.99$ (b) $\theta = 2$. Solid lines: simulation results; dashed lines: theoretical results.

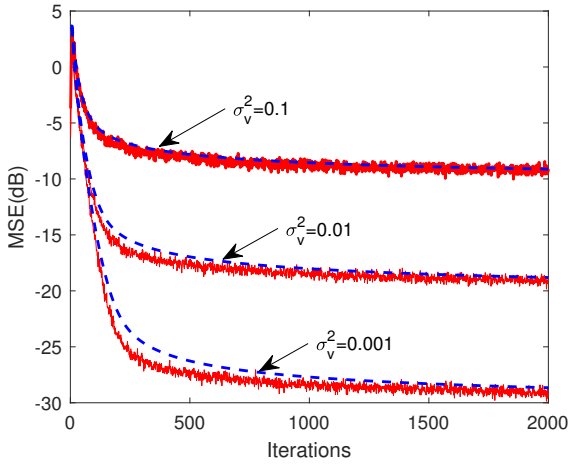


Fig. 9. MSE performance of the proposed algorithm against σ_v^2 for the AR1(0.5) input signal. $P = 1, M = 8, \alpha = 0.95$ and $\theta = 3$. Solid lines: simulation results; dashed lines: theoretical results.

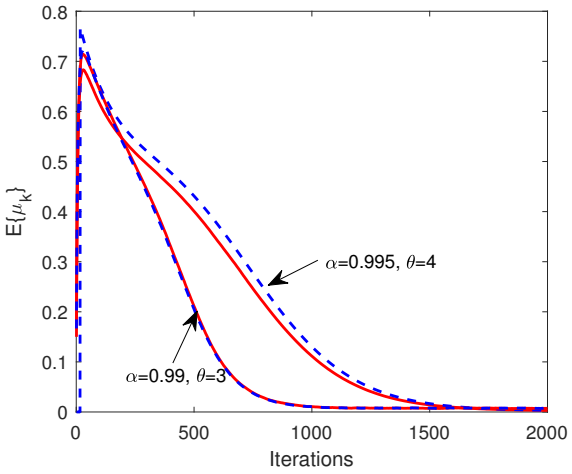


Fig. 10. Evolutions of the VSS for white input signal. $P = 1, M = 16$ and $\sigma_v^2 = 0.01$. Solid lines: simulation results; dashed lines: theoretical results.

transient MSE curves of the VSS-WLCNLMS algorithm. The theoretical results accurately predict the simulation results.

B. Wind Prediction

The sampled set of 24-h data was recorded at the Institute of Industrial Science, University of Tokyo [27]. The data set was collected using an ultrasonic anemometer and sampled at 50 Hz. In the experiments, we utilize the data set from the high wind speed region. The moving average filter is used to reduce the effects of high frequency noise. We consider the case of $P = 1$ for the proposed algorithm and AAPA to implement the one step ahead prediction, and the AAPA with $P = 1$ turns into the normalized version of the WL-CLMS, namely the WL-CNLMS algorithm.

The data containing north-south (V_N) and east-west (V_E) direction readings of wind speed, is used to create the complex

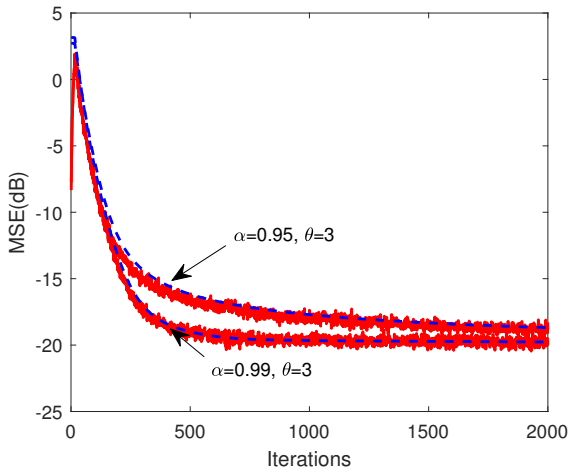


Fig. 11. MSE curves of the proposed VSS-WLCNLMS algorithm for white input signal. $P = 1, M = 16$ and $\sigma_v^2 = 0.01$. Solid lines: simulation results; dashed lines: theoretical results.

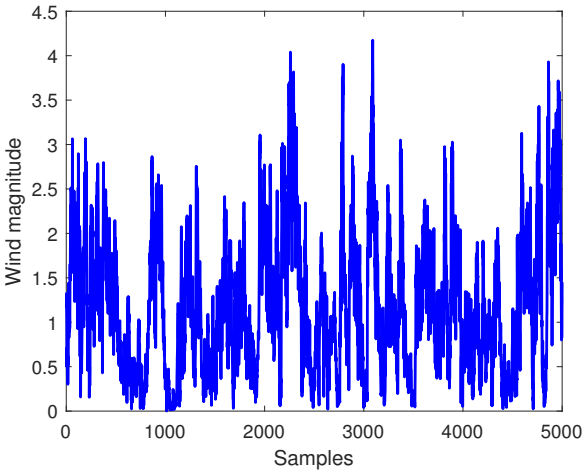


Fig. 12. Complex wind signal magnitude.

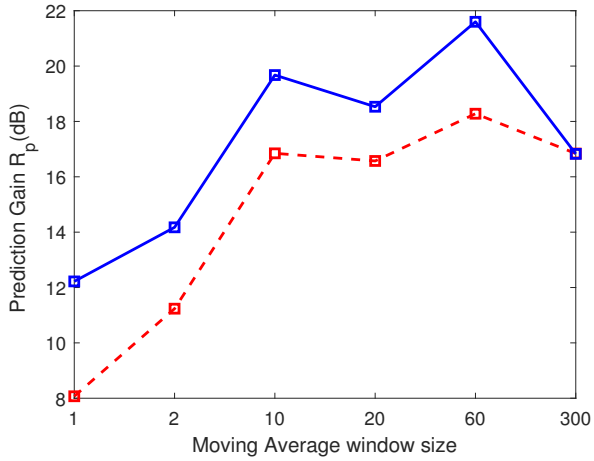


Fig. 13. Prediction gain of the WL-CNLMS (red line) and VSS-WLCNLMS (blue line) algorithms.

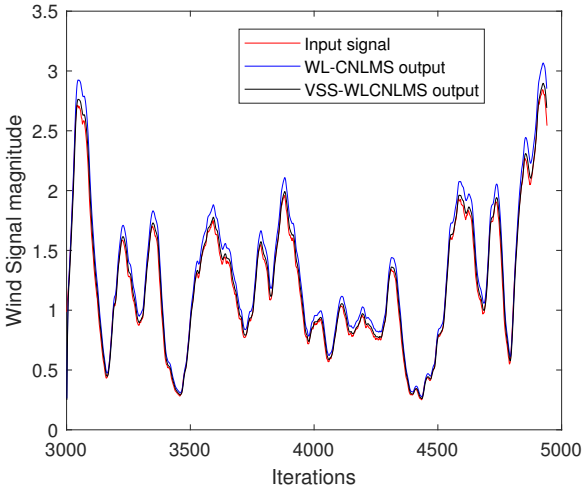


Fig. 14. Input and predicted signal using the WL-CNLMS and VSS-WLCNLMS algorithms.

wind signal V_c as follows:

$$\begin{aligned} v_{mag} &= \sqrt{V_E^2 + V_N^2} \\ \psi &= \arctan\left(\frac{V_N}{V_E}\right) \\ V_c &= v_{mag} e^{i\psi} \end{aligned} \tag{68}$$

The wind magnitudes are shown in Fig. 12. The first 3000 samples are used as the training set, and the last 2000 samples are used as the testing set. In this application, the parameters $M = 10, \alpha = 0.95, \theta = 3$, and $\gamma_s = 0.01$ are used for the proposed VSS-WLCNLMS. The step-size for the WL-CNLMS algorithm is set to $\mu = 0.01$.

Performance measure: The prediction gain R_p is used as the evaluation metric to measure the performance, given by [27]

$$R_p = 10\log 10\left(\frac{\sigma_{v_{mag}}^2}{\hat{\sigma}_e^2}\right) \tag{69}$$

where $\sigma_{v_{mag}}^2$ denotes the variance of the input signal $\{v_{mag}(k)\}$, and $\hat{\sigma}_e^2$ is the estimated variance of the forward prediction error $\{e(k)\}$. The sequence of $\{v_{mag}(k+1) - v_{mag}(k)\}$ is treated as the noise sequence to obtain the noise variance used in (21).

Fig. 13 shows the prediction gain of the VSS-WLCNLMS and WL-CNLMS algorithms with respect to the moving average window size. It is seen that the VSS-WLCNLMS algorithm shows better prediction gain than the WL-CNLMS algorithm. The prediction gain of the algorithms reaches a peak value for the window size 60. In the following experiment, the moving average window size is set to 60.

Fig. 14 shows the prediction results. As can be seen, the prediction using the WL-CNLMS algorithm has noticeable deviation from the input signal, while the VSS-WLCNLMS algorithm accurately predicts the input signal.

V. Conclusion

In this paper, we have proposed the VSS widely linear complex-valued APA (VSS-WLCAPA), which provides both the fast convergence and low steady-state misalignment. The VSS is derived based on minimizing the power of the augmented noise-free *a posteriori* error vector.

By exploiting the recursion for the covariance matrix of the weight error vector, we perform the theoretical analysis of the steady-state MSE of the proposed algorithm, where the dependency between the weight error vector and the noise vector is taken into account to improve the prediction accuracy. In the analysis, by employing the polar coordinate transformation, we derive the probability density function of the magnitude of the error to calculate the mean step-size.

Furthermore, we have studied the case when the projection order is set to one. By using the assumption of independence between the weight error vector and noise, we obtain the theoretical expressions for the transient and steady-state MSE of the algorithm. Simulation results for system identification scenarios have shown that the proposed algorithm outperforms the ACNSAF and AAPA. The theoretical results have been confirmed to provide good prediction of the algorithm behavior. Wind prediction experiments demonstrate the superiority of the proposed algorithm as well.

Appendix A
Calculation of $\mathbf{\Phi}_4$ and $\mathbf{\Phi}_5$

Substituting (35) into (33), we have

$$\begin{aligned}\mathbf{\Phi}_4 &= E[\tilde{\mathbf{w}}(k)\mathbf{v}^H(k)\mathbf{S}(k)\mathbf{U}^H(k)] = \\ &- E\Bigg\{\sum_{j=1}^{P}\Bigg(\prod_{p=1}^{j}\Big[\mathbf{I}-\mu_{k-p+1}\mathbf{U}(k-p+1)\mathbf{S}(k-p+1)\\ &\mathbf{U}^H(k-p+1)\Big]\Bigg)\mu_{k-j}\mathbf{U}(k-j)\mathbf{S}(k-j)\\ &\mathbf{v}(k-j)\mathbf{v}^H(k)\mathbf{S}(k)\mathbf{U}^H(k)\Bigg\}.\end{aligned} \tag{A.1}$$

Since the background noise $v(k)$ is noncircular white Gaussian and independent of the input signal, (A.1) can be further expressed as

$$\begin{aligned}\mathbf{\Phi}_4 &= \\ &- \sigma_v^2\sum_{j=1}^{P}E\Bigg\{\Bigg(\prod_{p=1}^{j}\Big[\mathbf{I}-\mu_{k-p+1}\mathbf{U}(k-p+1)\mathbf{S}(k-p+1)\\ &\mathbf{U}^H(k-p+1)\Big]\Bigg)\mu_{k-j}\mathbf{U}(k-j)\mathbf{S}(k-j)\mathbf{Z}^j\mathbf{S}(k)\mathbf{U}^H(k)\Bigg\},\end{aligned} \tag{A.2}$$

where $\mathbf{Z}$ is a $P\times P$ matrix:

$$\mathbf{Z}=\begin{bmatrix}\mathbf{0} & \mathbf{I}\\ 0 & \mathbf{0}\end{bmatrix}. \tag{A.3}$$

Similarly, $\mathbf{\Phi}_5$ is given by

$$\begin{aligned}\mathbf{\Phi}_5 &= \\ &- \sigma_v^2\sum_{j=1}^{P}E\Bigg\{\mathbf{U}(k)\mathbf{S}(k)\mathbf{U}^H(k)\\ &\Bigg(\prod_{p=1}^{j}\Big[\mathbf{I}-\mu_{k-p+1}\mathbf{U}(k-p+1)\mathbf{S}(k-p+1)\\ &\mathbf{U}^H(k-p+1)\Big]\Bigg)\mu_{k-j}\mathbf{U}(k-j)\mathbf{S}(k-j)\mathbf{Z}^j\mathbf{S}(k)\mathbf{U}^H(k)\Bigg\}.\end{aligned} \tag{A.4}$$

Appendix B
Validity of Approximation 2

To find the approximation for $E\left\{\frac{x}{y}\right\}$, we first define the mean values:

$$\mu_x = E(x), \quad \mu_y = E(y). \tag{B.1}$$

Using a Taylor series expansion of $\frac{x}{y}$ around μ_x, μ_y gives rise to

$$\begin{aligned}\frac{x}{y} &\approx \frac{x}{y}\Big|_{\mu_x,\mu_y}\\ &+ (x-\mu_x)\frac{\partial}{\partial x}\left(\frac{x}{y}\right)\Big|_{\mu_x,\mu_y} + (y-\mu_y)\frac{\partial}{\partial y}\left(\frac{x}{y}\right)\Big|_{\mu_x,\mu_y}\\ &+ \frac{(x-\mu_x)^2}{2}\frac{\partial^2}{\partial x^2}\left(\frac{x}{y}\right)\Big|_{\mu_x,\mu_y} + \frac{(y-\mu_y)^2}{2}\frac{\partial^2}{\partial y^2}\left(\frac{x}{y}\right)\Big|_{\mu_x,\mu_y}\\ &+ (x-\mu_x)(y-\mu_y)\frac{\partial^2}{\partial x\partial y}\left(\frac{x}{y}\right)\Big|_{\mu_x,\mu_y}\\ &+ o\left(\left((x-\mu_x)\frac{\partial}{\partial x}+(y-\mu_y)\frac{\partial}{\partial y}\right)^3\left(\frac{x}{y}\right)\right).\end{aligned} \tag{B.2}$$

By taking the expectation in (B.2), we arrive at the second order approximation (ignoring all terms higher than two)

$$\begin{aligned}E\left\{\frac{x}{y}\right\} &\approx \frac{\mu_x}{\mu_y} - \frac{\mathrm{cov}(x,y)}{\mu_y^2} + \mathrm{var}(y)\frac{\mu_x}{\mu_y^3}\\ &= \frac{\mu_x}{\mu_y}\left[1-\frac{\mathrm{cov}(x,y)}{\mu_x\mu_y}+\frac{\mathrm{var}(y)}{\mu_y^2}\right],\end{aligned} \tag{B.3}$$

where $\mathrm{cov}(x,y) = E(xy) - E(x)E(y)$ denotes the covariance between x and y, and $\mathrm{var}(y) = E(y^2) - E(y)^2$ is the variance of y [48], [49].

The quantities $\sigma_{\mathbf{e}}^2(k)$ and $\sigma_{\mathbf{e}_a}^2(k)$ in (19) and (20) can be written into the following recursive forms

$$\sigma_{\mathbf{e}}^2(k) = \sum_{j=0}^{k}\lambda_j\|\mathbf{e}(j)\|^2, \quad \sigma_{\mathbf{e}_a}^2(k) = \sum_{j=0}^{k}\lambda_j\|\hat{\mathbf{e}}_a(j)\|^2, \tag{B.4}$$

where $\lambda_j = (1-\alpha)\alpha^{k-j}$.

Let $x = \sigma_{\mathbf{e}_a}^2(k)$ and $y = \sigma_{\mathbf{e}}^2(k)$, then taking the expectation in (18) yields

$$E(\mu_k) = E\left\{\frac{\sum_{j=0}^{k}\lambda_j\|\hat{\mathbf{e}}_a(j)\|^2}{\sum_{j=0}^{k}\lambda_j\|\mathbf{e}(j)\|^2}\right\}. \tag{B.5}$$

In (B.3), if the absolute value of $-\frac{\mathrm{cov}(x,y)}{\mu_x\mu_y}+\frac{\mathrm{var}(y)}{\mu_y^2}$ is much smaller than one, i.e., $\left|-\frac{\mathrm{cov}(x,y)}{\mu_x\mu_y}+\frac{\mathrm{var}(y)}{\mu_y^2}\right| << 1$, we have $E\left\{\frac{x}{y}\right\} \approx \frac{E(x)}{E(y)}$. Since it is difficult to rigorously prove that $\left|-\frac{\mathrm{cov}(x,y)}{\mu_x\mu_y}+\frac{\mathrm{var}(y)}{\mu_y^2}\right| << 1$, we verify this inequality in simulation, as shown in Fig. 15. With the increase of iterations, $\left|-\frac{\mathrm{cov}(x,y)}{\mu_x\mu_y}+\frac{\mathrm{var}(y)}{\mu_y^2}\right|$ increases slowly, and approaches a constant in the steady-state. During the whole evolution, $\left|-\frac{\mathrm{cov}(x,y)}{\mu_x\mu_y}+\frac{\mathrm{var}(y)}{\mu_y^2}\right| << 1$ always holds. Therefore, the approximation $E\left\{\frac{x}{y}\right\} \approx \frac{E(x)}{E(y)}$ can be used in the analysis.

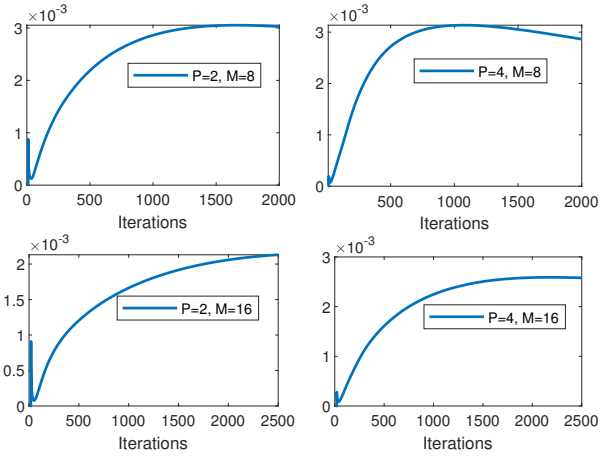


Fig. 15. Evolution of $\left|-\frac{\mathrm{cov}(x,y)}{\mu_x\mu_y}+\frac{\mathrm{var}(y)}{\mu_y^2}\right|$ for the AR1(0.95) input signal. $\alpha = 0.95, \theta = 1$ and $\sigma_v^2 = 0.01$.

APPENDIX C
DISTRIBUTION OF THE MAGNITUDE $|e(k)|$

A complex-valued Gaussian random variable $z = x + iy$ can be written in polar coordinates as:

$$z = r\exp(i\theta_p), \tag{C.1}$$

where $r = |z|$ denotes the magnitude of z, and θ_p is the phase of z. Suppose that its mean ρ and variance σ^2 are given by

$$E(z) = \rho = |\rho|\exp(i\gamma), \tag{C.2}$$

$$\mathrm{var}(z) = E|z-\rho|^2 = \sigma^2, \tag{C.3}$$

where γ is the phase of ρ.

According to the real multivariate Gaussian theory [50], [51], for an n-dimensional complex random vector $\mathbf{z}$ with mean vector $\boldsymbol{\rho}$ and positive definite covariance matrix $\boldsymbol{\Omega}$, that is, $E(\mathbf{z}) = \boldsymbol{\rho}$ and $E(\mathbf{z}-\boldsymbol{\rho})(\mathbf{z}-\boldsymbol{\rho})^H = \boldsymbol{\Omega}$, its probability density function (PDF) is given by [50]

$$f(\mathbf{z}) = \frac{1}{\pi^n \det\boldsymbol{\Omega}}\exp\left[-(\mathbf{z}-\boldsymbol{\rho})^H\boldsymbol{\Omega}^{-1}(\mathbf{z}-\boldsymbol{\rho})\right]. \tag{C.4}$$

When $n = 1$, we have

$$f(z) = \frac{1}{\pi\sigma^2}\exp\left(-\frac{|z-\rho|^2}{\sigma^2}\right). \tag{C.5}$$

The joint PDF $f(r,\theta_p)$ of r and θ_p is related to $f(z)$ by [50]

$$\begin{aligned} f(r,\theta_p) &= rf(z) \\ &= \frac{r}{\pi\sigma^2}\exp\left(-\frac{\left(r^2+|\rho|^2\right)}{\sigma^2}\right)\exp\left[\frac{2|\rho|r}{\sigma^2}\cos(\theta_p-\gamma)\right]. \end{aligned} \tag{C.6}$$

Using (C.6), the marginal density function $\varphi(r)$ of r is calculated as

$$\begin{aligned} \varphi(r) &= \int_0^{2\pi} f(r,\theta_p)\mathrm{d}\theta_p \\ &= \frac{r}{\pi\sigma^2}\exp\left(-\frac{\left(r^2+|\rho|^2\right)}{\sigma^2}\right)\int_0^{2\pi}\exp\left[\frac{2|\rho|r}{\sigma^2}\cos(\theta_p-\gamma)\right]\mathrm{d}\theta_p \\ &= \begin{cases} \frac{2r}{\sigma^2}\exp\left(-\frac{\left(r^2+|\rho|^2\right)}{\sigma^2}\right)I_0\left(\frac{2|\rho|r}{\sigma^2}\right), & r\geq 0 \\ 0, & r<0 \end{cases} \end{aligned} \tag{C.7}$$

where I_0 is the zero-order modified Bessel function of the first kind.

In particular, if z is zero-mean, then

$$\varphi(r) = \begin{cases} \frac{2r}{\sigma^2}\exp\left(-\frac{r^2}{\sigma^2}\right), & r\geq 0 \\ 0, & r<0 \end{cases} \tag{C.8}$$

Therefore, the magnitude $|e(k)|$ has the PDF given in (C.8).

ACKNOWLEDGMENT

The work of L. Shi and H. Zhao was partially supported by National Science Foundation of PR China (Grant: 61571374, 61871461, and 61433011), Sichuan Science and Technology Program (Grant: 19YYJC0681), National Rail Transportation Electrification and Automation Engineering Technology Research Center under Grant NEEC-2019-A02 and Doctoral Innovation Fund Program of Southwest Jiaotong University (Grant: D-CX201819). The work of Y. Zakharov was supported by the U.K. EPSRC (Grants EP/P017975/1 and EP/R003297/1).

REFERENCES

[1] A. H. Sayed, *Fundamentals of Adaptive Filtering*. John Wiley & Sons, 2003.
[2] S. D. Stearns, "Adaptive Signal Processing," 1985.
[3] S. S. Haykin, *Adaptive Filter Theory*. 4th ed: Prentice-Hall, 2002.
[4] J. Benesty and S. L. Gay, "An improved PNLMS algorithm," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, 2002, pp. II–1881.
[5] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, S. L. Gay *et al.*, *Advances in Network and Acoustic Echo Cancellation*. Springer, 2001.
[6] M. Rupp, "A family of adaptive filter algorithms with decorrelating properties," *IEEE Transactions on Signal Processing*, vol. 46, no. 3, pp. 771–775, 1998.
[7] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Electronics and Communications in Japan (Part I: Communications)*, vol. 67, no. 5, pp. 19–27, 1984.
[8] S. G. Sankaran and A. L. Beex, "Convergence behavior of affine projection algorithms," *IEEE Transactions on Signal Processing*, vol. 48, no. 4, pp. 1086–1096, 2000.
[9] H.-C. Shin, A. H. Sayed, and W.-J. Song, "Variable step-size NLMS and affine projection algorithms," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 132–135, 2004.

[10] C. Paleologu, J. Benesty, and S. Ciochina, "A variable step-size affine projection algorithm designed for acoustic echo cancellation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 8, pp. 1466–1478, 2008.

[11] L. R. Vega, H. Rey, and J. Benesty, "A robust variable step-size affine projection algorithm," *Signal Processing*, vol. 90, no. 9, pp. 2806–2810, 2010.

[12] S. Werner and P. S. Diniz, "Set-membership affine projection algorithm," *IEEE Signal Processing Letters*, vol. 8, no. 8, pp. 231–235, 2001.

[13] I. Yamada, K. Slavakis, and K. Yamada, "An efficient robust adaptive filtering algorithm based on parallel subgradient projection techniques," *IEEE Transactions on Signal Processing*, vol. 50, no. 5, pp. 1091–1101, 2002.

[14] Y. Zakharov and F. Albu, "Coordinate descent iterations in fast affine projection algorithm," *IEEE Signal Processing Letters*, vol. 12, no. 5, pp. 353–356, 2005.

[15] Y. V. Zakharov, "Low-complexity implementation of the affine projection algorithm," *IEEE Signal Processing Letters*, vol. 15, pp. 557–560, 2008.

[16] S.-E. Kim, S.-J. Kong, and W.-J. Song, "An affine projection algorithm with evolving order," *IEEE Signal Processing Letters*, vol. 16, no. 11, pp. 937–940, 2009.

[17] H.-C. Shin and A. H. Sayed, "Mean-square performance of a family of affine projection algorithms," *IEEE Transactions on Signal Processing*, vol. 52, no. 1, pp. 90–102, 2004.

[18] S. J. de Almeida, J. C. M. Bermudez, N. J. Bershad, and M. H. Costa, "A statistical analysis of the affine projection algorithm for unity step size and autoregressive inputs," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, no. 7, pp. 1394–1405, 2005.

[19] H. Rey, L. R. Vega, S. Tressens, and J. Benesty, "Variable explicit regularization in affine projection algorithm: Robustness issues and optimal choice," *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 2096–2109, 2007.

[20] S.-E. Kim, J.-W. Lee, and W.-J. Song, "A theory on the convergence behavior of the affine projection algorithm," *IEEE Transactions on Signal Processing*, vol. 59, no. 12, pp. 6233–6239, 2011.

[21] P. Park, C. H. Lee, and J. W. Ko, "Mean-square deviation analysis of affine projection algorithm," *IEEE Transactions on Signal Processing*, vol. 59, no. 12, pp. 5789–5799, 2011.

[22] T. K. Paul and T. Ogunfunmi, "On the convergence behavior of the affine projection algorithm for adaptive filters," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 58, no. 8, pp. 1813–1826, 2011.

[23] J. Shin, C. H. Lee, N. Kong, and P. Park, "An affine projection algorithm with update-interval selection," *IEEE Transactions on Signal Processing*, vol. 61, no. 18, pp. 4600–4609, 2013.

[24] M. Ferrer, A. Gonzalez, M. de Diego, and G. Pinero, "Distributed affine projection algorithm over acoustically coupled sensor networks," *IEEE Transactions on Signal Processing*, vol. 65, no. 24, pp. 6423–6434, 2017.

[25] Y. Xia, S. C. Douglas, and D. P. Mandic, "Adaptive frequency estimation in smart grid applications: Exploiting noncircularity and widely linear adaptive estimators," *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 44–54, 2012.

[26] L. Liu, R. Zhang, and K.-C. Chua, "Multi-antenna wireless powered communication with energy beamforming," *IEEE Transactions on Communications*, vol. 62, no. 12, pp. 4349–4361, 2014.

[27] D. Mandic, S. Javidi, S. L. Goh, A. Kuh, and K. Aihara, "Complex-valued prediction of wind profile using augmented complex statistics," *Renewable Energy*, vol. 34, no. 1, pp. 196–201, 2009.

[28] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE Transactions on Signal Processing*, vol. 43, no. 8, pp. 2030–2033, 1995.

[29] D. P. Mandic and V. S. L. Goh, *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*. John Wiley & Sons, 2009, vol. 59.

[30] S. Javidi, M. Pedzisz, S. L. Goh, and D. P. Mandic, "The augmented complex least mean square algorithm with application to adaptive prediction problems," in *IAPR Workshop Cognitive Information Processing Santorini, Greece*, 2008, pp. 54–57.

[31] Y. Xia and D. P. Mandic, "Widely linear adaptive frequency estimation of unbalanced three-phase power systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 1, pp. 74–83, 2011.

[32] Y.-M. Shi, L. Huang, C. Qian, and H.-C. So, "Shrinkage linear and widely linear complex-valued least mean squares algorithms for adaptive beamforming," *IEEE Transactions on Signal Processing*, vol. 63, no. 1, pp. 119–131, 2014.

[33] E. C. Mengüç and N. Acır, "An augmented complex-valued least-mean kurtosis algorithm for the filtering of noncircular signals," *IEEE Transactions on Signal Processing*, vol. 66, no. 2, pp. 438–448, 2017.

[34] Y. Xia, C. C. Took, and D. P. Mandic, "An augmented affine projection algorithm for the filtering of noncircular complex signals," *Signal Processing*, vol. 90, no. 6, pp. 1788–1799, 2010.

[35] P. Wen, J. Zhang, S. Zhang, and D. Li, "Augmented complex-valued normalized subband adaptive filter: Algorithm derivation and analysis," *Journal of the Franklin Institute*, vol. 356, no. 3, pp. 1604–1622, 2019.

[36] L. Shi, H. Zhao, and Y. Zakharov, "Performance analysis of shrinkage linear complex-valued lms algorithm," *IEEE Signal Processing Letters*, vol. 26, no. 8, pp. 1202–1206, 2019.

[37] Z. A. Bhotto and A. Antoniou, "A family of shrinkage adaptive-filtering algorithms," *IEEE Transactions on Signal Processing*, vol. 61, no. 7, pp. 1689–1697, 2012.

[38] M. A. Iqbal and S. L. Grant, "Novel variable step size NLMS algorithms for echo cancellation," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2008, pp. 241–244.

[39] H.-C. Huang and J. Lee, "A new variable step-size NLMS algorithm and its performance analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 2055–2060, 2011.

[40] L. Shi, H. Zhao, and Y. Zakharov, "Generalized variable step size continuous mixed p-norm adaptive filtering algorithm," *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2018.

[41] L. Shi, H. Zhao, W. Wang, and L. Lu, "Combined regularization parameter for normalized LMS algorithm and its performance analysis," *Signal Processing*, vol. 162, pp. 75–82, 2019.

[42] S. Zhang, W. X. Zheng, and J. Zhang, "A new combined-step-size normalized least mean square algorithm for cyclostationary inputs," *Signal Processing*, vol. 141, pp. 261–272, 2017.

[43] R. H. Kwong and E. W. Johnston, "A variable step size LMS algorithm," *IEEE Transactions on Signal Processing*, vol. 40, no. 7, pp. 1633–1642, 1992.

[44] S. Zhang and W. X. Zheng, "Mean-square analysis of multi-sampled multiband-structured subband filtering algorithm," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 66, no. 3, pp. 1051–1062, 2018.

[45] S. Koike, "A class of adaptive step-size control algorithms for adaptive filters," *IEEE Transactions on Signal Processing*, vol. 50, no. 6, pp. 1315–1326, 2002.

[46] V. J. Mathews and Z. Xie, "A stochastic gradient adaptive filter with gradient adaptive step size," *IEEE Transactions on Signal Processing*, vol. 41, no. 6, pp. 2075–2087, 1993.

[47] H.-S. Lee, S.-E. Kim, J.-W. Lee, and W.-J. Song, "A variable step-size diffusion LMS algorithm for distributed estimation," *IEEE Transactions on Signal Processing*, vol. 63, no. 7, pp. 1808–1820, 2015.

[48] J. Hayya, D. Armstrong, and N. Gressis, "A note on the ratio of two normally distributed variables," *Management Science*, vol. 21, no. 11, pp. 1338–1341, 1975.

[49] K. Wolter, *Introduction to Variance Estimation*. Springer Science & Business Media, 2007.

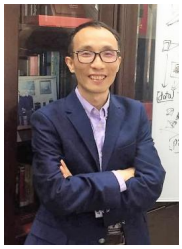
[50] K. S. Miller and K. S. Miller, *Complex Stochastic Processes: an Introduction to Theory and Application*. Addison-Wesley NewYork, 1974.

[51] A. Lapidoth, *A Foundation in Digital Communication*. Cambridge University Press, 2017.

Long Shi was born in Jiangsu Province, China, in 1991. He is currently pursuing a PH.D. degree in electrical engineering at Southwest Jiaotong University. He was a visiting student with University of York from 2018 to 2019. His interesting topics are adaptive signal processing, statistical signal processing, frequency estimation in power systems, channel estimation in communications, and Massive MIMO.

Haiquan Zhao (M'11-SM'17) was born in Henan Province, China, in 1974. He received the B.S. degree in applied mathematics in 1998, M.S. degree and Ph.D degree in signal and information processing all at Southwest Jiaotong University, Chengdu, China, in 2005 and 2011, respectively. Since 2012, he was a Professor with the School of Electrical Engineering, Southwest Jiaotong University, Chengdu, China. From 2015 to 2016, as a visiting scholar, he worked at University of Florida, USA. His current research interests include nonlinear active noise control, information theoretical learning, neural networks, adaptive network, adaptive filtering algorithm, and power system frequency estimation. At present, he is the author or coauthor of more than 140 international journal papers (SCI indexed), and the owner of 44 granted invention patents. Prof. Zhao has served as an active reviewer for several IEEE Transactions, IET series, Signal Processing, and other international journals. At presented, Prof. Zhao is a handling editor of Signal Processing, and also Associate Editor of IEEE ACCESS, and AEU- International Journal of Electronics and Communications.

Yuriy Zakharov (M'01 - SM'08) received the M.Sc. and Ph.D. degrees in electrical engineering from the Power Engineering Institute, Moscow, Russia, in 1977 and 1983, respectively. From 1977 to 1983, he was with the Special Design Agency in the Moscow Power Engineering Institute. From 1983 to 1999, he was with the N. N. Andreev Acoustics Institute, Moscow. From 1994 to 1999, he was with Nortel as a DSP Group Leader. Since 1999, he has been with the Communications Research Group, University of York, UK, where he is currently a Reader in the Department of Electronic Engineering. His research interests include signal processing, communications, and underwater acoustics.



Badong Chen was born in Sichuan Province, China. He received the B.S. and M.S. degrees in control theory and engineering from Chongqing University, in 1997 and 2003, respectively, and the Ph.D. degree in computer science and technology from Tsinghua University in 2008. He was a Post-Doctoral Researcher with Tsinghua University from 2008 to 2010, and a Post-Doctoral Associate at the University of Florida Computational NeuroEngineering Laboratory (CNEL) during the period October, 2010 to September, 2012. He visited the Nanyang Technological University (NTU) as a visiting research scientist during July to August 2015. He is currently a professor at the Institute of Artifificial Intelligence and Robotics (IAIR), Xi'an Jiaotong University. His research interests are in signal processing, information theory, machine learning, and their applications in cognitive science and the engineering.

Yaoru Yang obtained his Bachelor of Engineering degree in Communication Engineering in 2010, Master of science in Digital Signal Processing in 2015 and his Ph.D in Electronic Engineering in 2019. After receiving his Bachelor degree in 2010, he worked as an engineer till he began his Msc study in the UK. He is interested in Statistical Signal Processing and Computational Neuroscience, including adaptive signal processing algorithm, spectrum analysis in neuronal recordings and synchronisation analysis in spike train data.